User Activity Aware Support System Using Activity Frame

Nicholas Melo and Jaeryoung Lee, *Member, IEEE*

***Abstract*— This work presents a system is able to support the users on their daily livings using an activity and intention recognition method. The system is designed to be focused on the applicability, working in real time. The recognition method uses the concept of activity frame, which is defined as a set of sequenced environmental observations containing meaningful information (such as objects' locations, sensors' activation, etc) related to the recognition of activities and tasks accomplished in one location. Analyzing the specific frame, it is possible to relate, through a set of conditions, the observed states to a specific activity or intention. By analyzing the frequency of those activities and intentions occurrences, it is possible to identify unusual behavior and guide an smart interactive device, such as robot, to support the user. The proposed recognition method was tested with the data provided by an smart home project, and the recognition rate for the proposed method has high accuracy, based on other similar ones. The information of activities intentions can provide meaningful guidelines for the robot.**

I. INTRODUCTION

The work presented in this paper is developed as part of the MIC and EC Horizon 2020 project CARESSES that designed the cultural-aware robots for elderly care [1]. In recent years, there has been an increasing in the usage of applications which uses a wide range of sensors in the houses or health-care facilities. The price of the sensors, along with easy-to-use micro controllers have contributed to that trend [2]. Equipping users' residences with the sensors, it is possible to collect a wide variety of user related information that can be used by several applications, such as security, home automation and health-care [2], [3], [4]. The general idea behind those applications is to extract context-aware information in order to improve specific home services or identify problems regarding the user quality of life, and recognizing health issues in early stages[5].

Some issues regarding the quality of life can be identified analyzing several aspects related to how the user is accomplishing their daily life activities. For example, for an elderly living alone, if the time needed to finish the cooking activity is increasing compared to previous days, then that time increment could be a indicator related to a loss of skill required to finish the task. Furthermore, integrating the activity recognition and analysis system with a humanoid robotic device, can lead to development and improvement of human-robotic interaction systems. These activity aware robot-supported systems can help the user directly to accomplish the tasks in their daily live activities as considering their cultural and personal backgrounds. Thus, a key factor for an activity aware support system is the integration between the activity recognition module and the real world interactive component, in the form of a robot.

Usually, activity recognition (AR) systems are composed by different elements, each one related to a specific function, such as activities semantic definition, data storage, context network and recognition engine [4], [6]. Most of the recent AR approaches usually use conditional probability as a tool to find the coherent assignment of sensor triggers that are related to a specific activity. For example, some systems uses dynamic optimization to find the most likely sequence of user-environment related conditions, which is able to be linked to a given set of activities using a engine based on Hidden Markov Model [7]. The sequences of sensors activation are inferred hidden states in the model.

Because of the limited information given by most of simple sensors, a wide range of activities recognition approaches relies on using visual information [8], [9], and the sequences of the complex data are used to identify specific user body motion, postures and parameters, which can be related to a set of activities [10], [8], [11]. However, due to privacy concerns inside the users houses or care facilities, the use of visual information is still a controversial topic [9], [12]. Moreover, this restriction, for some cases, leaded to the need of AR system that won't use camera or image as resources, instead system that uses a set of simple sensors to recognize activities has been increasing in the past years [9].

Focusing on the implementation aspects of AR systems, one downside of probabilistic based approaches is the representation of the hidden states used in the context, which can lead to strange observations. The out-of-context activities inferences can lead to recognition of unnatural behaviors, for example preparing soup without water, taking the phone without placing a call or taking a bath while working on the computer [9], [13].

Also, usually the estimation method requires some computing processing power when dealing with a high degree of conditions and number of sensors. The possible delay caused by complexes approaches can be a drawback for a real time robotic application focusing on interacting with the user while its activities are recognized by a AR module [4]. This problem can be solved by specialized hardware and extensive training stages, however, it is not feasible for applications where fast, dynamic and low cost implementations have a high demand.

Nicholas Melo and Jaeryoung Lee are with the Chubu University, Kasugai, Japan nicholas.bastos@gmail.com; jaeryounglee@isc.chubu.ac.jp

In this work, we propose an applicability driven architecture of an activity aware support system. A novel activity recognition system presents to be focused on a real time implementation using one of nods in human robot interaction, in order to support the user on their daily activities. The proposed recognition method organize the sensory related information as environmental state instances that can be sequenced in a time interval called activity frame. Analyzing this specific frame, the activities of users can be recognized according to a conditional matrix. Using a time interval to analyze the sensory data, it is less likely that the system will fail to recognize specific activities due to momentary loss of communication between sensors[4]. Moreover, the activity instance is divided into core of the activity and activity intention. Those two elements can be used not only to provide a better understanding of the activity in the range of the temporal constrains, but also to deliver more support options and references to the robotic element.

II. User Activity and Intention

Activity can be defined as a set of actions, aiming to a specific objective, accomplished in a certain period of time, in a choose or specified location. Focusing on daily living activities, the objective is to accomplish basic functions needed by the human to maintain a standard quality of life, such as cooking, taking a bath, eating, and so on. For recognizing of basic daily life activities in a house environment without visual information, most of the activities can be related to a tool (or set of tools) and a location.

Observations regarding the user location along with activation status of a specific tool or object can be used to recognize an activity. For example, for the activity, *cooking*, the user needs to use at least one tool (a bowl for example) in one place (kitchen). The combination and classification levels of those two elements, location and tool, can be defined by an ontology in the form of descriptions logic [11], [13], [9].

A. User activity

The semantic development regarding the description and classification of activities and their related triggers events are not the focus of this works, instead, it is used the semantic description of activities proposed by [9]. In order to recognize the activities, we define as the $environment\_status(t)_j$ the combination of all the information that can be extracted from the location j in the time stamp t. An activity instance is defined by a specific set of multiple instances of $environment\_status(t)_j$. This definition will be used by the proposed recognition process.

B. User Intention

Because of the high variability regarding the location and time for the activity to be executed, it is no trivial task to associate the precise instance of a specific activity. For example, for the activity *bathing*, it is possible to relate the start of the activity with the moment where the shower sensor is triggered. However, should be interesting for an smart environment warm up the water while the user is preparing for the activity *bathing*. Hence, whenever the user enters the bathroom the activity intention for bathing can be initialized, even without directly sensory trigger from the main tool (shower). This aspect is indirectly related to definition of the activity and for the instance trigger conditions as well.

For that reason, in the present work, we split the structure of activities into two elements, the core of the activity and the activity intention. The core of the activity is directly related to an sensor activation, mainly the tool, or set of tools, used for the activity. The activity intention is identified by an range of $environment\_status(t)_j$ that usually happens before the core of the activity. For example, for the activity *watching tv*, the core of the activity starts when the user turn the television on, while the activity intention start whenever the user come closer to the TV or the remote control. Using the idea of activity intention, the problem of ambiguous inference can be approached in a different manner, allowing to new possibilities regarding the user interaction with other smart devices present in the home environment. The inclusion of the concept of core of the activity and activity intention for the ontology level is now under development.

III. Activity Aware Support System

The main objective of the Activity Aware Support System is to track in real time the user activities using a recognition method, and with that information create guidelines and interaction references for the robot to interact with the user, supporting it in their daily tasks. Both the activities characteristics and interaction actions should be set a prior by a knowledge base. The overview of the proposed method is show in the Fig. 1.

First, sensory data is collected from the environment. The distribution of the sensors depends on the context and the list of activities to be recognized. It is important that each sensor node send information constantly according to a sampling time, and not only when the sensor is activated. This prevent failure of recognition performance in case where the sensor its triggered and the transmitted information is lost due to bad connectivity.

How each sensory information will be collected and synchronized it is not important for the proposed method. Due to recent advances in the sensory and micro controller area, the applicability of sensor network has been increasing, allowing the sensory system to have a wide range of possibles configurations for the sensor layer.

The second module, the data collector layer, validates raw data and stores it in the database. Also, this layer can process more complex sensory information in order to store environmental status. For example, the position of a certain object can be found by the calculation of data sent by different sensors.

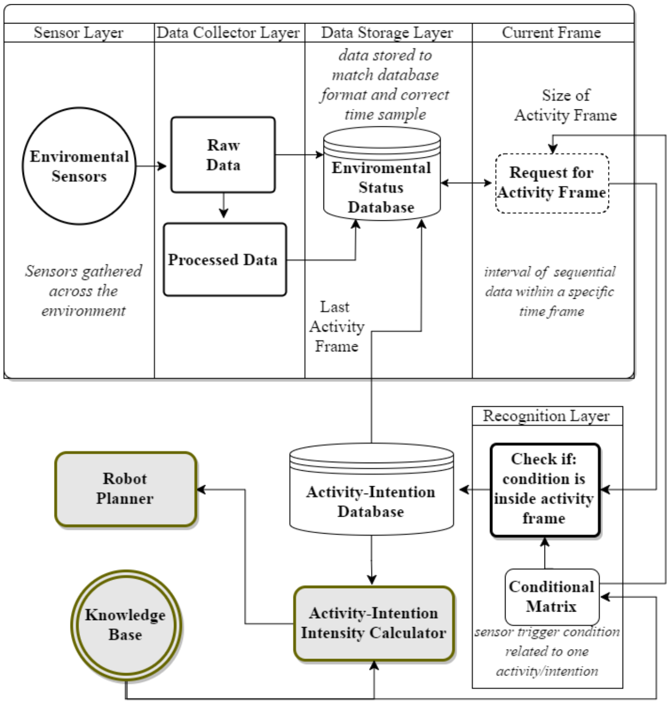

Fig. 1. Overview for the proposed method.

The next module, the data storage layer should store information regarding the environment and all the components inside it. We define $environment\_status(t)_j$ as the combination of the elements current states inside the location j, such as presence sensor status, user and object current position, tool usage data, in a specific time t. This information, in a real time application, is constantly updated for $environment\_status(t+1)_j$. According to the used network of sensor defined by the data collector layer, the update rate can be dynamically changed. The structure of the environmental status database is show in the Fig.2.

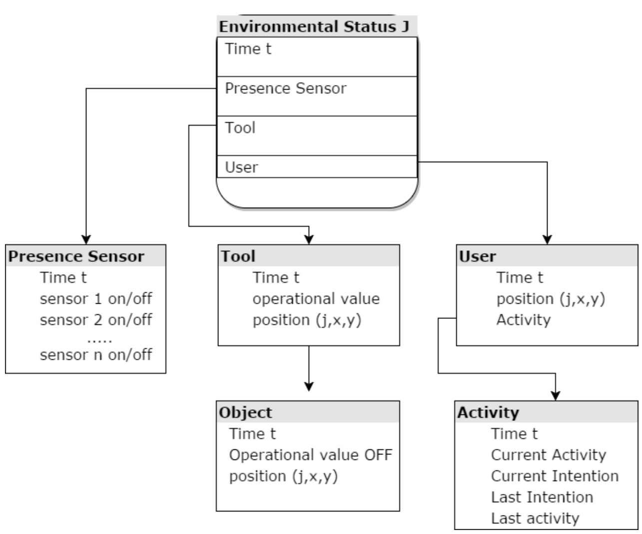

Fig. 2. Structure of the Environmental Status Database.

The environmental status is composed by the instance of *presence sensor*, *tool*, and *user* classes in a given time t. The *presence sensor* class has the information regarding all proximity sensor activation's present in the location j. The *tool* class, has the attributes operational value and position. For *object*, which can be identified as a sub-class of *tool*, all the operational values are *OFF*. The operational values can be either Boolean values (on or off), or real numbers. For example, the operational value of the tool *sink* can represent the water usage for that time, while the same attribute for the tool *TV*, can be *ON* or *OFF*.

The *user* class contains information about the user state. If feasible for the application, the user body related information can also represented for this class, for example, user body temperature, voice frequency and so on. For a simple implementation, the *user* class should show its position in the environment j and its current and last activity frame.

The current user-context environmental status is then analyzed in the form of an activity frame, which is used by the recognition layer. The objective of these two modules is to identify in real time the current user activity and intention through analysis of the present environmental status condition.

A. Activity Frame and Conditional Matrix

We define the activity frame as a time interval containing environmental information, in the form of multiples instances of $environment\_status(t)_j$ as show in the Eq. 1, where *ES* is the environment_status. An example of the activity frames is shown in the Fig. 3.

$$activity\_frame = [ES(frame_{start})_j, \ldots, ES(frame_{end})_j] \tag{1}$$

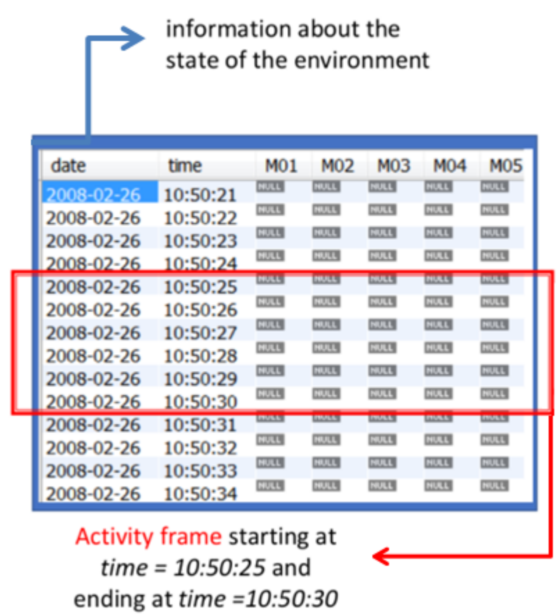

Fig. 3. Example of activity frame.

The size of the activity frame can vary according to each activity and it is defined by the knowledge base or conditional matrix. This value should be closest to the minimum value required to identify the usage of the sensors used to recognize the activity. The activity frame size used for the present work was 5 seconds.

Each activity frame is then compared to a conditional matrix which contains the requirements and conditions regarding the state of the environment (sensor data and object operational status) to trigger a specific activity/intention. Each row in the conditional matrix is related to one activity, and each column represents a set of environmental status conditions needed for the correct recognition of the activities core and activity intention. The system is capable of adding requirement of sensor activation sequence in the conditional matrix, in this case, the process present in the recognition layer will also look for a specific sequence sensor trigger. An example of core of the activity and activity intention condition is show in the Fig. 4.

Core Activity Condition:

activity_1 = wash hands;

core_activity_condition_1 = water usage sensor ON
AND
sink proximity sensor ON;

Intention Condition:

activity_1 = wash hands;

activity_intention_condition_1 = sink proximity sensor ON;

Fig. 4. Example of Activity Conditions

B. Data Processing

If the condition is found within the activity frame, then the core of the activity or activity intention is set as recognized. After the comparison, the activity frame is moved forward. The activity frame selection and comparison process rate works in parallel to the sensor layer. The process data flow for the recognition processed used can been seen in the Fig. 5.

The recognized activities/intentions are stored in an proper database. This activity-intention database is be used to update the current *user* class, but can also be used by external analysis modules, aiming in identifying the user habits or personal characteristics for example.

The activity-intention intensity calculator verify if the current activity or intention has usual frequency values. Basically the objective of this module is to identify strange duration for the activities and intentions and warn the planner. If a strange duration regarding one specific activity or intention is is observed, then the robot planner is activated in order to deal with this unusual behavior.

For example, if the system is noticing the duration of activity intention for *phone call* is increasing and there

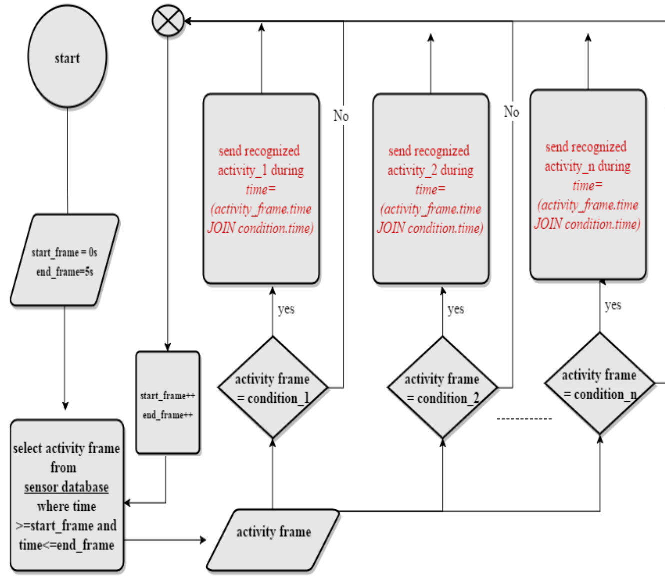

Fig. 5. Data flow

is no core of the activity detected, then the user is looking for the phone but can't find it. Because inside the environmental status database there is the information regarding the object position, the robot planner can be activated in order to inform where the object is, or get the object for the user.

The configuration of the usual duration for activity and intention is set by the knowledge base, which also set values for the conditional matrix. This is intimately related to the context where the user is performing the tasks and the activities themselves. For example, in a case where the core of the activity duration for *phone call* is increasing, there is no need for the robot to interfere.

IV. Experimental Results

It is important for the applicability of the proposed system to have a recognition method with high accuracy and easy to integrate with real time data acquisition system. To test the proposed recognition method, based on the concept of the activity frame, it was used data provided by the WSU smart home project [14]. These data sets represent sensor events collected in the WSU smart apartment and serves to meet research needs around testing of the technologies using real data through the use of a smart homes environment [14].

It was used the data collected from 10 different days, for a total of 20 different users. The number of experiments accomplished in one day varies between one and three per day. A total of five ADL activities were performed in sequence during the experiment, but for the present work it was used the activities (*make a phone call*) and (*cook*), identified as Activity 1 and Activity 2 consecutively. These activities were selected to validate the proposed method because they have direct relation to the concept of user intention and core of the activity.

It was added for every time value t present in the activity-intention database the activity label field, which represents the activity manually set by a annotator who watched the experiments.

Because the annotator was informed to label one activity per experiment in a determined sequence, the aspect of multitask was not tested in the present work.

Comparing the recognized activity and recognized intention values with the activity label, it is possible to calculate the recognition rate. Hence the recognition rate is directly relate when the activities core and activity intention were identified while the activity was labeled. The opposite also reduce the recognition rate, i.e., when the activity is identified but there is no activity label for that specific time value. The table Tab. I show the recognition rate for all days.

Day	Activity 1 Recognition Rate (%)	Activity 2 Recognition Rate (%)
2008/2/26	88.31	78.43
2008/2/27	98.36	86.25
2008/2/29	90.71	84.98
2008/3/3	94.11	74.52
2008/3/6	95.65	71.77
2008/3/7	85.61	87.61
2008/3/24	80.73	82.10
2008/3/26	98.16	70.25
2008/3/27	93.33	82.34
2008/4/15	99.01	92.21
All Days	**92.40**	**81.04**

TABLE I

RECOGNITION RATE FOR ALL DAYS

The recognition rate present in [15] shown a recognition rate of 58.9%, while using a ontology based method it was presented 72.30%, bot for the activity 1 (*make a phone call*). Although it is not know if the same data was used, and according to the limitation of the present experimental tests, it is possible to assume the presented method was able to recognize the activity in a similar high accuracy level.

The overall distribution between all days, with maximum, minimum, and average recognition rate can be seen in the form of a box plot in the Fig. 6 bellow.

The activity labeling process performed by the annotators during the experiments took the whole activity into consideration. However, in the present work it was used the concept of activity intention. Hence, as a said before, the activity flag by the annotator includes both the core of the activity and activity intention recognized by the presented method. It is possible to see this separation in the Fig. 7, where it shown the core of the activity and activity intention recognized time interval. For this specific activity, *make a phone call*, the core of the activity is recognize when the phone sensor is activated within the activity frame, while the activity intention is recognized whenever the user is closer to the phone. In this particular context, the phone position is fixed, although this could be changed for other scenarios as well.

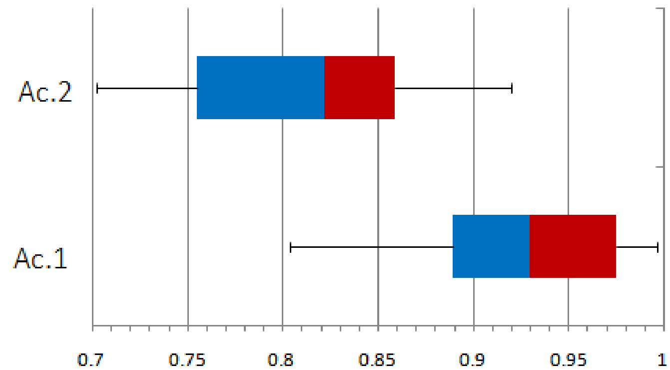

Fig. 6. Distribution for the Recognition Rate Between all Days

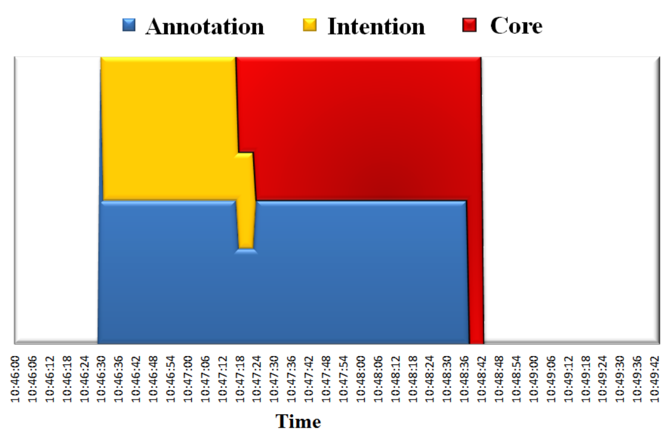

Fig. 7. Example of recognition results at one day for Activity 1

As the user get close to the phone location, the system recognize the related activity intention. The activity-intention intensity calculator keep the value of the sequenced frames while the activity intention is been identified. During the simulations using the experimental data from the dataset, it was noticed a need for a core of the activity post time parameter for some cases. This parameter guides the system to disable the recognition for the activity intention after the end of the core of the activity is recognized.

For example, after the user end the activity *make a phone call*, he still will be close to the phone, hence, the system will keep identifying the current status as activity intention. However, this happens only because the user need to move away from the phone after the activity ended. Using the core of the activity post time set as few seconds, it is possible to avoid this mistake. Further analysis need to be done regarding the ideal duration of the core of the activity post time. The second activity, *cook*, for its nature, provide different results and insights about the system behavior towards this activity type. The *cook* activity is more complex than the *make a phone call*, and can be divided in several sub-activities. However, the fundamental concept validated in this work is the same; a location and a tool (or set of tools) used to identify the core of the activity and activity intention. The intention is identified whenever the user is in the kitchen, with at least two objects used to cook outside their place of storage. The core of the activity is identified while the user uses the burner. From Figure 8 it is possible to see a different spread of recognition between activity intention and core of the activity when compared to the activity *make a phone call*.

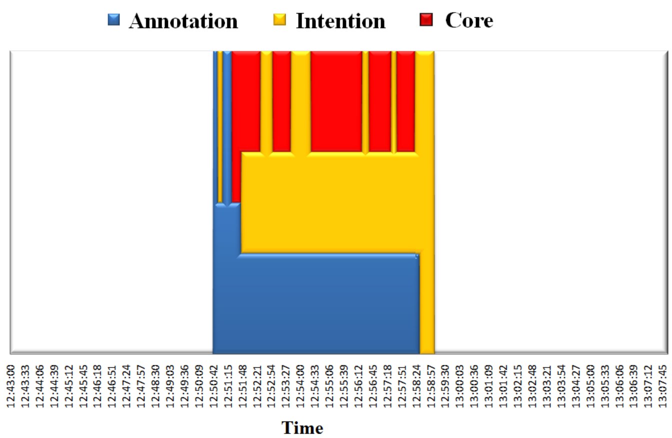

Fig. 8. Example of recognition results at one day for Activity 2

In a case where the activity intention for cook is increasing, but there is no core of the activity detected, the system, using the robot, tries to help the user finding one of the tools used to cook, or pointing out a problem in the burner. In this case the core of the activity post time is set to zero, once the user can turn on and off the main tool used to identify the core of the activity during performing the entire task. Analyzing the number of core of the activity identified in the same time period while activity intention was also been identified, can provide information about the user habits or conditions to perform the task.

V. Conclusions

This work presented a system able to recognize simple daily life activities in real time and use the observed information in order to guide a robotic system to support the user on maintaining or accomplish the tasks. Using the concept of activity frame and activity intention, allow the robotic planner more options to reference the robotic actions to support the user. The method is limited, at the moment, to only simple activities, with one subject at the same time, and it was tested with two daily activities. However, the results using the proposed recognition method, based on the activity frame, showed a similar high accuracy rate, compared to the state-of-art approaches. As a next step we intend to expand the concept of several elements used in the present work (such as activity frame, core of the activity and activity intention) in a ontology format, using description logic, and test the system in a real-time human robot interaction.

Acknowledgment

This project was supported by the EU-Japan coordinated R&D project on "Culture Aware Robots and Environmental Sensor Systems for Elderly Support" commissioned by the Ministry of Internal Affairs and Communications of Japan and EC Horizon 2020.

References

[1] B. Bruno *et al.*, "Paving the way for culturally competent robots: A position paper," in *26th IEEE International Symposium on Robot and Human Interactive Communication*, 2017, accepted.

[2] A. Perišić, M. Lazić, B. Perišić, and R. Obradović, "A smart house environment - the system of systems approach to model driven simulation of building (house) attributes," in *2015 IEEE 1st International Workshop on Consumer Electronics (CE WS)*, March 2015, pp. 56–59.

[3] S. Oh, W. Woo, *et al.*, "Camar: Context-aware mobile augmented reality in smart space," *Proc. of IWUVR*, vol. 9, pp. 48–51, 2009.

[4] E. M. Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," in *International Conference on Pervasive Computing*. Springer, 2004, pp. 158–175.

[5] P. N. Dawadi, D. J. Cook, and M. Schmitter-Edgecombe, "Automated cognitive health assessment using smart home monitoring of complex tasks," *IEEE transactions on systems, man, and cybernetics: systems*, vol. 43, no. 6, pp. 1302–1313, 2013.

[6] M. Buranarach, Y. M. Thein, and T. Supnithi, "A community-driven approach to development of an ontology-based application management framework," in *Joint International Semantic Technology Conference*. Springer, 2012, pp. 306–312.

[7] B.-C. Cheng, Y.-A. Tsai, G.-T. Liao, and E.-S. Byeon, "Hmm machine learning and inference for activities of daily living recognition," *The Journal of Supercomputing*, vol. 54, no. 1, pp. 29–42, 2010.

[8] R. Bodor, B. Jackson, and N. Papanikolopoulos, "Vision-based human tracking and activity recognition," in *Proc. of the 11th Mediterranean Conf. on Control and Automation*, vol. 1. Citeseer, 2003.

[9] D. Riboni, T. Sztyler, G. Civitarese, and H. Stuckenschmidt, "Unsupervised recognition of interleaved activities of daily living through ontological and probabilistic reasoning," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 1–12.

[10] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body worn inertial sensors," *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, p. 33, 2014.

[11] K. Wongpatikaseree, M. Ikeda, M. Buranarach, T. Supnithi, A. O. Lim, and Y. Tan, "Activity recognition using context-aware infrastructure ontology in smart home domain," in *Knowledge, Information and Creativity Support Systems (KICSS), 2012 Seventh International Conference on*. IEEE, 2012, pp. 50–57.

[12] A. GhaffarianHoseini, N. D. Dahlan, U. Berardi, A. GhaffarianHoseini, and N. Makaremi, "The essence of future smart houses: From embedding ict to adapting to sustainability principles," *Renewable and Sustainable Energy Reviews*, vol. 24, pp. 593–607, 2013.

[13] K. Wongpatikaseree, A. O. Lim, and Y. Tan, "A context-aware information in smart home for health recommendation service based on care architecture," in *Proceeding of the 2nd Asian Conference Information System*, 2013, pp. 501–508.

[14] D. J. Cook and M. Schmitter-Edgecombe, "Assessing the quality of activities in a smart environment," *Methods of information in medicine*, vol. 48, no. 5, p. 480, 2009.

[15] G. Singla, D. J. Cook, and M. Schmitter-Edgecombe, "Tracking activities in complex settings using smart environment technologies," *International journal of biosciences, psychiatry, and technology (IJBSPT)*, vol. 1, no. 1, p. 25, 2009.